KATHOLIEKE UNIVERSITEIT
LEUVEN

Faculty of Economics and
Applied Economics

Department of Economics

Legitimately Diverse, yet Comparable: on Synthesising Social Inclusion Performance in the EU

by

Laurens CHERCHYE
Wim MOESEN
Tom VAN PUYENBROECK

Public Economics

Center for Economic Studies
Discussions Paper Series (DPS) 03.01
http://www.econ.kuleuven.be/ces/discussionpapers/default.htm

January 2003

DISCUSSION
PAPER

Legitimately diverse, yet comparable: On synthesising social inclusion performance in the EU.

Laurens Cherchye[a,b,†]

Wim Moesen[b]

Tom Van Puyenbroeck[a,c,‡]

2003

Abstract

The Open Method of Coordination (OMC) intends to enhance EU member states' performance with regard to social inclusion. In this context a set of commonly agreed performance indicators plays an important role. While the communicative power of a synthetic indicator has been recognised, several objections have been raised against such a construction. In this paper, we argue that a set of separate indicators can in principle be combined into one synthetic performance index without giving up on the notion of subsidiarity, and without fundamentally impairing the peer pressure incentives that constitute an important rationale for OMC. We complement the presentation of the conceptual framework with a number of empirical applications, thereby indicating how the basic method may be instrumental for policy benchmarking practice.

[a] Faculteit Toegepaste Economische Wetenschappen, Katholieke Universiteit Leuven, Afdeling Kortrijk.

[b] Centrum voor Economische Studiën, Katholieke Universiteit Leuven

[†] Postdoctoral Fellow of the Scientific Research Fund – Flanders (FWO Vlaanderen)

[c] Centrum voor Openbare Financiën en Sociale Economie, Europese Hogeschool Brussel.

[‡] Corresponding author. Europese Hogeschool Brussel, Stormstraat 2, 1000 Brussel, Belgium. E-mail: tom.vanpuyenbroeck@prof.ehsal.be

Legitimately diverse, yet comparable: On synthesising social inclusion performance in the EU.

Laurens Cherchye

Wim Moesen

Tom Van Puyenbroeck

1. INTRODUCTION

There is a broad consensus throughout Europe that, even within an EMU, each nation represents a welfare state *sui generis.* It explains why member states are willing to remain responsible for their social policy, even if they explicitly acknowledge that they share many core social policy objectives. As phrased by Scharpf (2002a,b), institutional differences, distinct policy legacies and normative orientations imply that 'Europe' must preserve its own democratic legitimacy by explicitly taking the *legitimate diversity* of member states into account. These convictions motivate the EU's current key operational tool for enhancing member states' performance with regard to social inclusion: a modus operandi known as the Open Method of Coordination (OMC). This method was established and conceptually refined in 2000 at the Lisbon, Feira, and Nice European Councils.[1] Essentially, OMC is a procedure in which member states (biennially) provide National Action Plans on Social Inclusion (NAPincls). These link EU-wide objectives to the domestic situation and to (sub)national policy measures. Next, the NAPincls go through a process of peer reviewed monitoring, comparison and evaluation. The NAPincls also underlie the 'Joint Report on Social Inclusion' delivered by the Commission.

The wording *National* Action Plans is crucial, since it underscores that the effective choice of welfare policy measures is still the prerogative of national welfare states. The peer review of the NAPincls is expected to provide the impetus for national governments to improve upon the outcomes of their welfare policies. OMC has been categorized as a 'soft law' approach. This notion simultaneously conveys (i) the limited desirability and ability of direct EU steering of national policy making in the field of social policy and (ii) the idea that nation states *may* adapt their policies even in the absence of legally binding supranational steering instruments.

Given its recent inception, we do not know yet whether and how nation states may adapt their own welfare policies in the light of OMC. Observers have discerned at least three interrelated ways for OMC to effectively foster social inclusion and the combat against poverty. First, OMC is considered an intelligent, context-sensitive 'learning-by-monitoring'-procedure. That is, member states may learn from identified best practices in other countries and apply these,

[1] In fact, the generic notion OMC is nowadays applied in different policy areas. 'The' OMC we discuss in the paper, i.e. OMC as applied to social inclusion, is for example pre-dated by a somewhat dissimilar (and 'stronger'; see below) OMC for national employment policies. Its use (in stronger or weaker forms) will also be extended to other areas, such as e.g. pension policies.

mutatis mutandis, at home (see e.g. Hemerijck, 2002). For some sceptics, this begs the thorny question whether any member state has an incentive to adapt its own 'distinct' policy. The answer is related to a second alleged advantage: since OMC builds on the recognition of core social values via commonly agreed objectives, it should enable a more precise substantial definition of the somewhat fuzzy 'European Social Model'. And this can in turn be "a powerful driver to improve the quality of social protection in Europe" (Vandenbroucke, 2002). One may wonder at this point how 'the substance' can in practice become 'a powerful driver'. It seems that one needs the third avenue, as explicitly discussed by some authors, to answer the previous questions. Publicly scrutinized comparative benchmarking *may* eventually entail political pressure on national states. A 'virtuous competition' (Giammusso and Tangorra, 2002) could emerge with regard to social inclusion, as each member state is compared with others on account of the commonly defined objectives.

Using the conditional in many of the previous sentences is warranted. As the OMC is currently conceived, there remains a clear danger that this soft law tactic degenerates into spineless red-tapeism. Its success primarily hinges upon "the willingness of those which de facto control national policies" and, absent such willingness, NAPincls "may simply reflect business as usual and the unfortunate liaison officers attending innumerable rounds of meetings in Brussels may take the blame for national policies on which they have no influence." (Scharpf, 2002a, p. 33). More precisely, key decision-makers today face few incentives to adapt their welfare policies but those resulting from the OMC-comparative exercise *as (and a fortiori: if) it is received domestically.*[2] The last point is worth stressing in view of the criticisms raised by some direct stakeholders following the first actual implementation of the OMC on poverty and social exclusion. The European Anti-Poverty Network, for example, explicitly mentions that "national governments confront no penalties for non-compliance with their Plan". It further commented that there was almost no public awareness, no media interest and even little national parliamentary awareness about the first NAPincls, that these first NAPincls largely seem to have been written as reports for 'Europe' rather than tools for effective policy integration tools, and that "the lack of incentives to participate seriously may permanently consign the NAPincl process to the margins." (Duffy, 2002; see also Atkinson, 2002b).

From any of the above perspectives, the construction of a set of common indicators for social inclusion is an issue of utmost importance for the success of OMC. Its mere existence is a *conditio sine qua non* for going beyond the point of paying political lip service to the commitment at Lisbon of making a decisive impact on the eradication of poverty and social exclusion by 2010. Learning from each other can barely be done if a shared assessment toolkit is lacking. Moreover, the set of common indicators *de facto* embodies the commonly agreed objectives (Atkinson, 2002a; Vandenbroucke, 2002). They are therefore vital –for any stakeholder– for benchmarking, for measuring mutual and relative progress, etc. Notably for that reason, they should be considered as a key ingredient in providing the necessary domestic incentives for national policy makers.

As an upshot of the efforts by the Belgian Presidency in the second half of 2001, an agreement on a first set was reached at the Laeken Summit. It comprises 10 primary indicators which, together with secondary and tertiary indicators, are said to reflect the

[2] Contrary to the original OMC, which is used for employment policy, the common objectives on social inclusion do currently not translate into commonly shared *policy targets*, valid for each member state and scrutinized by the Commission. Another important difference is the lack of explicit financial tools in the social inclusion case, whereas the OMC-employment is linked to the European Structural Funds.

multidimensional nature of social inclusion.[3,4] It is important to note that the indicators all measure outputs: they are, intentionally, *performance indicators*. The choice to ignore inputs in the benchmarking exercise is again grounded in the concern to maintain the locus of control for social policy at the member states' level. The list is not definitive as other indicators will probably be included. Even in its current configuration one faces problems of data availability and comparability between member states. Also, and perhaps strangely enough, it only partially coincides today with the shorter list of "structural indicators for social cohesion" presented yearly by the Commission to allow for an objective assessment of the progress made towards the Lisbon Council objectives. Insofar as such sources of confusion are transitory, we will not address them here.

Rather, we start from the earlier recognition that commonly agreed indicators on poverty and social inclusion may be useful for many purposes, but that these underlying rationales seem to require different kinds of 'sets' of indicators. This holds especially as regards their dimensionality. On the one hand, from the pure 'learning'-perspective, it seems that a set of indicators can hardly be too small: there are many (intertwined) ways in which social exclusion manifests itself, and probably many best practices are available. On the other hand, the same set can quickly be regarded as too big in view of its capacity as an effective, easy-to-communicate means of advancing (domestic) public interest, i.e. as a leverage tool that implicitly aims at generating appropriate incentives for decision makers.

With these considerations as a background, we outline in this paper a method for constructing a *synthetic indicator* for social inclusion performance. We present this method in the next section after listing the main arguments that have been raised in favour or against a synthetic indicator. The central feature of the index we put forward is that, absent any consensus on social policy *priorities*, it grants each member state extensive leeway for deciding how to evaluate its own indicators. Importantly, when compared to the use of a set as a multidimensional benchmarking instrument *qua* incentive-generating tool, the obtained summary index does not lead to fundamentally different responses by member states. We provide an illustrative empirical application in section 3, demonstrate how to use the method for intertemporal analysis, and address some issues pertaining to sensitivity analysis. We also briefly touch upon the relationship between the synthetic (output) performance measure and the input side of the social policy equation. Section 4 offers some concluding remarks.

---

[3] A great deal of the preparatory work to the eventual establishment of a set at Laeken was carried out by A. Atkinson, B. Cantillon, E. Marlier and B. Nolan. Their carefully justified recommendations have meanwhile been published. The Belgian Presidency invited comments on the Atkinson et al.-report. Most of these comments, often related to the question whether the indicators are indeed adequate to capture social exclusion. These comments have been gathered in a special issue of *Politica Economica* (see the list of references at the end).

[4] Note, though, that several of the primary indicators are also reported via breakdowns by age, gender, or other characteristics. It leads Atkinson (2002) to state that "…some 70 numbers are requested as primary indicators, whereas fewer than half of that number are requested as secondary indicators. This is the reverse of the relationship we had in mind." Indeed, the concern for transparency provided a major justification for the recommendation in Atkinson et. al. (2002) that there should be a restricted number of primary indicators.

2. A SYNTHETIC INDICATOR FOR SOCIAL INCLUSION

2.1 Few pros and many cons?

To be sure, we are well aware that today many decision makers and experts are quite concerned about the emergence of league tables and the 'naming and shaming' which these could entail. Even if we recognise the value of incentives, we do share some concerns with them. As already indicated, 'one number' is a far too shallow basis for learning about the multiple factors causing poverty and social exclusion. Also, as it stands now there are undeniably a number of measurement problems regarding the indicators. We take it here that most of such data problems are transitory. And we underline that this problem is genuine to the (inter-country comparison of) individual components, not to a synthetic indicator *per se.* We here therefore only will survey the most important remarks that focus on the idea of merging individual components in one number, and take up other data issues in sections 3 and 4.

The (few) stated pros of doing so are all related to the idea of communication, the 'eye-catching property', i.e. fostering public awareness by 'providing the big picture' so as to avoid the fuzziness and possibly even the loss of credibility associated with a plethora of single indicators (see e.g. Atkinson et al., 2002; Hills, 2002; Joint Research Centre of the European Commission (JRC), 2002). We furthermore consider the remarks of Micklewright (2001) highly relevant: lacking a good synthetic index, there are always the dangers that (i) one reverts to constructing very crude country-specific indices, e.g. by averaging ranks, or by counting the number of national indicators below resp. above the European average; (ii) that none of the indicators gets sufficient attention; or even (iii) that excessive public attention is again focused on just one dimension, thus abolishing the original desideratum of respecting the multidimensional nature of social exclusion. Last but not least, it is certainly worthy to note that the Commission itself, within the framework of its structural indicators, is not unsympathetic to the idea of 'composite indicators'. It has already developed two such indices to help assessing the knowledge-based economy, and is willing to extend their use to other areas (COM (2002) 551 Final, 16.10.2002).

Yet, on balance, the idea of a synthetic indicator *for social exclusion* has hitherto bumped into criticism. A (too) simple big picture may send misleading messages and may invite one to draw simplistic policy conclusions (JRC, 2002). The multidimensionality of social exclusion also manifests itself in the fact that different indicators have no common measurement unit, which makes transformation to a common metric "a tricky operation" (Brandolini, 2002). In this particular respect, Atkinson et al. (2002) list several objections against the idea of using *rank* orders, both for the constituent individual indicators and for a synthetic indicator, *inter alia* because these do not convey how far away countries are from each other or from (external) targets. Additionally, they state, that "politically, the process will not encourage member states to learn from each other if attention is focused on a single rank order."

The main opposition is however related to the question whether and how to weigh the sub-indicators. Two matters of concern can be discerned. First, on a conceptual level, if one agrees on a set of (leading) indicators, this in itself may be taken as evidence that one has selected equally imperative components of social inclusion. The very idea of mutually weighting them has therefore been deemed undesirable by some authors: allowing for a trade-off between different dimensions is blocked on *a priori* grounds.[5] Logically, such a position entails that countries are only demonstrably outperformed if they are weakly dominated in *all* dimensions of the indicator-set by at least one other country. Only a lexicographic ordering of observations is allowed. More often than not this could result in the finding that the large majority of countries constitute 'their own benchmarks' when considering the full picture: they can just pick, choose, and point at sub-indicators to convey the message that they are indeed distinct, and not weakly dominated by any other country. Table 1 provides a simple hypothetical example: with just two indicators for 10 countries, there is nowhere a weak dominance relationship, and hence there are no overall grounds for any country to be considered outperformed on social inclusion if the two dimensions are judged equally imperative. While admittedly fictitious (some EU countries would in practice be outperformed by others in terms of Table 1's dimensions), it is easy to see that in reality the big picture quickly risks to become totally obscure as more indicators are added.

Table 1: a hypothetical example

Country	I	II	III	IV	V	VI	VII	VIII	IX	X
Poverty rate	10	9	7	12	18	19.9	20	6.9	5	4.9
Long term Unemployment rate	6	9	10	5.9	4	2.1	2	14	15	19

As a matter of fact, strictly adhering to the idea that all indicators are equally commanding in a global benchmarking exercise quite counter-intuitively implies that any country is effectively better off by consolidating its idiosyncracy. This need not be bad *per se*. But for genuine benchmarking exercises, or for virtuous competition, it clearly risks to get the above conception of 'not weighting' the indicators self-defeating.

Furthermore, it is hard to deny that scarce resources *de facto* compel governments to make choices and set priorities: any policy choice has its opportunity cost. Any national government –at least implicitly– in many instances has to consider how to mutually trade-off various policy outcomes. In fact, as regards the problem addressed in this paper, the related ideas of legitimate diversity and subsidiarity seem to suggest strongly that countries are even entitled to set different outcome priorities.

It is precisely the latter point which has lead to a second type of objections against the idea of weighting. These are 'practical', i.e. political objections. Again quoting Atkinson et al., "in the context of the EU, there are evident difficulties in reaching agreement on such weights, given that each member state has its own national specificity." It is this objection that we put central in the rest of our paper. Specifically, we take it that a good synthetic indicator for social inclusion should grant considerable discretion to each country when assigning its own policy weights, without allowing them to put all weight exclusively on one dimension. In the next subsection we demonstrate how this can be achieved.

[5] See e.g. Brandolini (2002): " For the sake of simplicity, -but the observation carries over to more complicated formulations- suppose that the summary index equals the arithmetic mean of the selected indicators. In adopting such an index, we are implicitly assuming that one unit more of indicator A can be substituted for one unit less of indicator B or vice versa. If A is the unemployment rate and B the proportion of people failing to reach 65, our summary index would suggest that the valuation of the social situation is unchanged when the unemployment rate is reduced by 1 when the unemployment rate is reduced by 1 percentage point at the same time as the proportion of people dying before 65 is raised by 1 percentage point. I do not think that this conclusion is acceptable, nor is it likely to gain wide acceptance."

2.2 A 'Subsidiarity-Respecting' Synthetic Indicator for Social Inclusion

Consider the general case of *m* indicators of social inclusion for *n* countries, and let y_{ij} be the value of indicator *i* in country *j*. We want to merge these individual indicators into a single-valued synthetic indicator, defined as the weighted average of the original set of *m* single-dimensional indicators. As we eventually want to evaluate 'policy' performance, the weights should preferably reflect policy priorities. Within that perspective, however, it is indeed difficult to specify *a priori* any generally acceptable weights to be accorded to each indicator. Therefore, we adopt a weighting procedure that is based on the principle of 'benefit-of-the-doubt'.

This procedure was originally proposed in the context of macroeconomic performance evaluation by Melyn and Moesen (1991); see also Cherchye (2001) for a methodological discussion. Since then, similar methods have been applied to other types of macro-level assessments. For example, Mahlberg and Obersteiner (2001) and Cherchye and Kuosmanen (2002) adopted benefit-of-the-doubt weighting for cross-country assessments of human development and sustainable development performance.

A study that is more closely related to ours is that of Storrie and Bjurek (2000), who analyze the labour market performance of EU countries. Our analysis is slightly different in scope. In addition, and more importantly, we adopt a rather different methodological perspective. Specifically, while Storrie and Bjurek build their argument on characterizing the set of performance benchmarks (as a convex 'efficiency frontier' of the observed vectors), our procedure explicitly starts from weighting the individual social inclusion indicators (and selects benchmarks directly within the set of *observed* countries). While both approaches are related in a mathematical sense (cf. the duality theorem of linear programming), the ('primal') weighting orientation we adopt seems particularly attractive in the context of 'easy to convey' performance evaluations.

NORMALIZATION

Generally, the construction of a synthetic indicator proceeds in two steps. In a first step the original data, i.e. the country-specific values of the sub-indicators, are normalized. We note here that some social inclusion indicators can be considered as 'bads' (i.e. higher values represent worse performance; e.g., long term unemployment rate) while others can be considered as 'goods' (i.e. higher values represent better performance; e.g., life expectancy). To render goods and bads commensurable, any normalization procedure for social inclusion indicators should take this distinction into account. We denote the normalized counterpart of each y_{ij} by y_{ij}^{n}. To simplify notation, we further use $y_i^{\min}$ to refer to the lowest value for the *i*-th indicator over all countries in the sample, and $y_i^{\max}$ to refer to the highest value. The particular method we use in this paper takes the difference between the sub-indicator value of country *j* and the corresponding value of the worst-performing country, and divides this difference by the observed sample range. Thus, when the *i*-th indicator is a good, we have:

$$y_{ij}^{n} = \frac{y_{ij} - y_i^{\min}}{y_i^{\max} - y_i^{\min}}.$$

And alternatively, if the *i*-th indicator is a bad, we have:

$$y_{ij}^{n} = \frac{y_i^{\max} - y_{ij}}{y_i^{\max} - y_i^{\min}}$$

In both cases, the values of the normalized indicators vary between 0 and 1, 0 always corresponding to the worst (cross-section) performance in the sample and 1 corresponding to the best performance. And, importantly in view of the criticisms raised against mere rankings, the normalized figures clearly reflect the intensity of the relative performance differences.

A slightly modified approach that deserves special mentioning in the present context is one where actual performance is evaluated with respect to an *exogenously determined benchmark value* for the sub-indicators. In the above formulas, the reference value for each indicator is endogenously selected as the best practice within the sample of countries, which is in line with the current gist of political arguments. Yet, the possibility of setting explicit targets for a social inclusion indicator, or of comparing with best practices in other, non-EU countries, can be readily incorporated in the procedure. For example, the same normalization procedure, with exogenous benchmarks $Y_i^{\min}$ and $Y_i^{\max}$ replacing our $y_i^{\min}$ and $y_i^{\max}$, is used for the construction of the Human Development Index (HDI; see UNDP, 2001).

Indeed, the same or highly similar normalization procedures underlie many other existing synthetic indicators. Still, it is worth to point out at this point that the weighting method discussed below is also applicable to alternative normalizations (see, e.g., JRC (2002) for a recent assessment of alternatives).

Actually, we can expect the normalization procedure to have a rather limited impact on the global performance results. In particular, it can be shown that a normalization procedure only affects the results of our synthetic performance measure through restrictions that are eventually imposed on the implicitly selected policy weights, which reflect the relative importance of each social inclusion performance dimension (see below).[6] Stated otherwise, (only) if one wants to impose such weight restrictions, then the different indicators need to be expressed in a comparable measurement unit. In our opinion, imposing weight restrictions should be possible, and the normalization procedure we accordingly presented (which allows for a straightforward incorporation of exogenous benchmarks) is especially appropriate.

AGGREGATION

After normalization, the next step typically involves aggregation, per country, of the single-dimensional performance indices into a weighted sum. The central issue here pertains to the specification of the weights. Aggregation/weighting questions have been extensively studied in the literature on productivity indices; see, e.g., Balk (2002) for a discussion. The synthetic social inclusion index that we propose below is firmly rooted in that literature.

---

[6] Specifically, it can be shown that the performance assessment model without weight restrictions (see (2)) is a special version of the Charnes et al. (1978) Data Envelopment Analysis (DEA) model. Absent any weight restrictions (but non-negativity), a normalization procedure is then not even required. See, e.g., Lovell and Pastor (1995) for invariance properties of DEA models with respect to measurement units.

Normalized indicators can in principle be aggregated in several ways. In the most simple case one uses predetermined weights associated with each single indicator. For example, one can take the average of the different single indicators, which implies equal weighting. Another example, which deviates only slightly from the equal weighting procedure, is the construction of the HDI (UNDP, 2002). However, as discussed before, we believe that the indicator weights should be sensitive to national policy priorities. In that respect, it seems hardly tenable that each country *de facto* assigns equal weights to the different dimensions of social inclusion. As noted before, the specification of *a priori* weights seems a very difficult task when differing policy priorities prevail over countries.

By contrast, the procedure adopted in this paper *reconstructs* 'implicit' (or 'shadow') policy weights *from the observed performance* (i.e. *a posteriori*). More specifically, the relative weight accorded to each dimension is *endogenously* determined in our performance evaluation model, and reflects the associated relative performance of the country under evaluation. Stated otherwise, good relative performance in a particular dimension is seen as 'revealed' evidence of setting high national policy priority to that dimension. This seems an attractive second best route in the absence of full information about the true policy priorities.

The proposed methodology selects the most favorable weights for each country. Attractively, the resulting performance index hence respects the 'legitimate diversity'-aspect of social inclusion policies. In particular, policy makers can hardly claim that an unfair weighting scheme is employed for evaluating their country as any other weight profile would only worsen the position *vis-à-vis* the other countries in the sample. The method reveals optimal priority orderings for each evaluated country conditional upon the observed performance in each dimension.

Notwithstanding the constructive treatment of countries' diversity, another appealing feature of the method is its flexibility: it still allows for imposing various kinds of additional weight restrictions. Indeed, while it is hardly conceivable that experts will ever agree on 'point estimates' for country-specific weights accorded for each policy dimension, it seems much more reasonable to assume that they can reach consensus on bounds to be respected by the relative policy weights. Such consensus positions are readily incorporated in the proposed weighting procedure, as we discuss further on.

Modeling a synthetic Index for Social Inclusion

Generally, we can define a synthetic index of social inclusion in the OMC-context as the ratio of a country's actual performance over its 'benchmark performance', where performance is measured as a weighted sum of the single indicator values. Benchmark performance is represented by the (normalized) benchmark vector y^* of social inclusion indicators. Using w_i ($i = 1,...,m$) to represent the weight accorded to each indicator i and y_i^* ($i = 1,...,m$; $y^* = \left(y_1^*,...,y_m^*\right)$) to indicate benchmark performance in each policy dimension i, we get for each country j ($j = 1,...,n$):

$$\frac{\sum_{i=1}^{m} y_{ij}^n w_i}{\sum_{i=1}^{m} y_i^* w_i}.$$

Two questions remain in making this relative performance measure operational: (1) we need to define benchmark performance, and (2) we need to specify the relative weights accorded to the different dimensions of social inclusion.

The first question is easy to solve. We identify best practice from observed performance; we simply select the country which maximizes the overall performance value under the weights w_i ($i = 1,...,m$) as the benchmark, i.e.:

$$y^* = \arg \max_{y_k : k\in\{1,...,n\}} \left(\sum_{i=1}^{m} y_{ik}^n w_i \right).$$

Note that, since the weights are country-specific, nothing precludes that each country is compared with its proper best practice; other countries can be associated with different benchmark observations.

Our solution to the second question has been informally introduced above: we select the weights that maximize the resulting performance values for the country (j) *under evaluation*. At this stage, we only impose that the weights cannot be negative, i.e. that the synthetic value is a non-decreasing function of the individual social inclusion indicators. This gives the following performance measure for each country j (where *SI* stands for 'Synthetic Index' and 'Social Inclusion'):

$$SI_j = \max_{w_i \geq 0, i=1,...,m} \frac{\sum_{i=1}^{m} y_{ij}^n w_i}{\max_{y_k : k\in\{1,...,n\}} \sum_{i=1}^{m} y_{ik}^n w_i}. \qquad (1)$$

Clearly, as $\sum_{i=1}^{m} y_{ij}^n w_i \leq \max_{y_k : k\in\{1,...,n\}} \sum_{i=1}^{m} y_{ik}^n w_i$ for each weighting scheme w_i ($i = 1,...,m$), we have $0 \leq SI_j \leq 1$; and higher index values can be interpreted as better overall performance. The generosity of the method is immediate: the weights are chosen in such a way that no other weight combination would yield a higher relative performance value for country j. And if SI_j < 1, there are other countries $k \in \{1,..,n\} \setminus \{j\}$ in the rest of the sample that would obtain a higher overall index value, *even when using country j's most favorable weights*.

We illustrate the approach, applying it to our earlier example, in Table 2. Now we find that some countries are demonstrably outperformed by others in terms of *SI*, even when applying their most favorable weights. The *SI* differences are not very pronounced, which reflects benefit-of-the-doubt weighting. Additionally, we recall that the procedure we adopted is sensitive to the intensity of performance differences. This feature here most notably shows up in country VI's *SI*-value, which is extremely close to that of country VII (cf. their underlying original values in Table 1). Of course, in reality, differences could well be more pronounced as will be shown in our applications in section 3.

To illustrate the impact of a benefit-of-the-doubt perspective, we compare the *SI* results with those of a measure obtained by a priori imposing equal weights on the two single performance dimensions; see again Table 2, where '*EW*' stands for '*E*qual *W*eights'.[7] Unsurprisingly, the *SI* values are never below the *EW* values, thus directly demonstrating the impact of endogenously selecting country-specific most favorable policy weights. In some cases the differences are rather pronounced. Two notable examples in that respect are countries VII and X, which both achieve an *SI* value of 100%, while their *EW* value only amounts to about 70%. These two countries are heavily specialized in one particular policy dimension: they mainly focus their policy on either poverty abatement (country X) or unemployment abatement (country VII); see also the original data in Table 1. Such specialization effects show up in the implicit weights that underlie the computation of the respective *SI* values. We return to this specialization issue later, when we will discuss the restrictions on policy weights.

Table 2: synthetic indicators for the hypothetical example

Country	I	II	III	IV	V	VI	VII	VIII	IX	X
SI Value	100%	93.67%	100%	95.88%	92.94%	99.65%	100%	90.96%	100%	100%
EW Value	100%	92.27%	97.43%	91.13%	71.12%	70.13%	70.08%	81.41%	86.10%	70.08%

The social inclusion index can also be expressed in a linear programming form. To see this, note that only relative weights matter; multiplying all weights w_i ($i = 1,...,m$) by a common factor will not alter the index value. We can consequently normalize the weights so that $\max_{y_k : k\in\{1,...,n\}} \sum_{i=1}^{m} y_{ik}^{n} w_i = 1$, and thus re-express expression (1) as:

$$SI_j = \max_{wi} \sum_{i=1}^{m} y_{ij}^{n} w_i \ ; \qquad (2)$$

$$\text{s.t.}$$

$$\sum_{i=1}^{m} y_{ik}^{n} w_i \leq 1 \qquad \forall k = 1,...,n \ ;$$

$$w_i \geq 0 \qquad \forall i = 1,...,m \ .$$

Hence, standard linear programming software packages suffice to compute the aggregated social inclusion index, which is evidently an appealing feature for practical applications. We note that the outcomes of such exercise for each country j ($j = 1,...,n$) do not only yield the associated relative performance value, but also estimates for the (country-specific, dimension-specific) implicit policy weights.

As alluded upon before, the above model is readily complemented with additional weight restrictions. For example, in our basic empirical application we will use the following (and rather minimal) restriction:

$$w_i \geq \frac{1}{10}\left(\sum_{k=1}^{m} w_k\right) \qquad \forall i = 1,...,m \ . \qquad (3)$$

This restriction imposes that any social inclusion dimension that is considered in the model should get an (implicit policy) weight of at least 10%. In the same vein, upper bound

[7] For each country, we obtain this *EW* measure by taking the (equally weighted) average of the normalized values for the two performance indicators, and consequently dividing this value by the maximal average value in the sample (in casu that corresponding to country I).

restrictions can be added. For example, to impose that any dimension can at most get half of total policy weights, one would add:

$$w_i \leq \frac{1}{2}\left(\sum_{k=1}^{m} w_k\right) \qquad \forall i = 1,\ldots,m. \tag{4}$$

Observe that these restrictions are linear in the weights, which makes that adding either or both to the above programming problem does not change the linear nature of that problem. Of course, more specific (even country-specific) weight restrictions that reflect prior information (e.g. on expert consensus positions) can further refine the basic evaluation model.[8] In view of the above discussion, it is however evident that such refinements will, *ceteris paribus*, never lead to a higher synthetic index value for a country.

We conclude our hypothetical example by calculating such weight-restricted *SI* values, viz. by adding restriction (3) to the basic model. The obtained performance values do not deviate much from those associated with the non-restricted measure. The corresponding correlation coefficient amounts to almost 99%: the SI values for most countries remain (quasi) the same. The only sizeable change is associated with country X, which now achieves an *SI* value of 98.09%. We noted earlier that this was a 'specialized' country. Although still rather marginal, country X's lower *SI* value so illustrates that weight restrictions can reduce the 'adverse' impact of extreme specialization in a particular policy dimension, which apparently comes at an excessive cost in terms of the other performance dimensions. Our application to EU countries will reveal additional (and more convincing) examples of this specialization-correction effect.

2.3 The index versus the scoreboard

When constructing a summary indicator for social inclusion per country, the method just discussed seems most apt to comply with the demanding issue of legitimate diversity. Another advantage of the method is that its basic principles –weights are assigned so as to take care of a country's relative performance in the most generous way– are easily conveyed to the general public. They are hard to defy by national decision makers as any other weighting scheme would only generate a lower their index value for their country. Such features are desirable for an 'eye-catching' and 'incentive generating' metric.

It is also crucial to point out that, relative to the case in which the nowadays existing ('disaggregated') set would in fact be considered by stakeholders as a genuine *score*board, the proposed index does not fundamentally alter the nature of incentive-driven reactions by national decision makers. For one thing, if a country realizes a relative performance improvement somewhere on the scoreboard, this would directly be reflected in an increased index-value (remember that the normalization procedure preserves intensity differences with a common reference). Strictly speaking, this observation does not hold for the best overall performer(s), since by construction there is a maximum value of 1 for the index. One should however realize that the improvement would then turn up via a widened gap with the second-best performer. Moreover, the reason for fixing an upper bound echoes our option for streamlining *relative*, i.e. within-sample benchmarking. Stated differently, the only rationale

---

[8] Compare with Pedraja-Chapparo (1997), who discuss the issue of imposing weight restrictions within the context of DEA, and Cherchye and Kuosmanen (2002) who discuss of alternative types of weight restrictions that can be used for macro-level performance assessments.

for claiming that "the best in class" has additional room for improvement can be found in *external* benchmarks (i.e. adding other countries to the sample) or exogenously imposed targets, which, as we have shown above, can readily be incorporated at the normalization stage. Again, exactly the same remark holds for the scoreboard as such. With regard to this issue, we additionally stress that the method can also be applied to panel data, so that *intertemporal* progress or regress can equally be reported by the summary figure. We return to this remark in section 3.2.

Finally, from one point of view, the best (i.e. politically least contestable) summary index may seem to be the one which is exactly based on the set of commonly agreed indicators. Yet, if we want the rules of the benchmarking exercise to be fully similar for all member states, there can only be an exact correspondence between the existing set and the synthetic index if sufficient, reliable data are available for each dimension and each country (i.e. if there are no missing data entries in the scoreboard). As we have indicated, today this presents a practical difficulty. At first glance, this may be taken as the Achilles' heel of the entire exercise. But once more, this observation in itself does not tilt the balance in favour of an (incomplete) n x m-table. Quite the reverse, there are some good reasons to advocate a synthetic index in this case, even if that index only inexactly corresponds to the original, incomplete set. We take up this issue in the following section, where empirical applications will be examined.

3. Empirical applications

3.1 Basic Application

We first present an application of model (2) with only 4 performance dimensions included, viz. the low income rate, the income quintile ratio, the long-term unemployment rate and early school leavers (the precise definitions –still in use December 1, 2002– can be found in Annex I of the 'Joint Report on Social Inclusion' (2002))[9]. These indicators appear both on the current (OMC) scoreboard for social inclusion and on the Commission's list of 'structural indicators for social cohesion'. The associated data are tabulated in Appendix 1. Effectively, this implies that we start by neglecting 6 primary indicators of the current OMC-scoreboard (and 3 structural indicators), the reason being that not all 15 EU countries provide data for these indicators yet.

Whether this marks today's (practical) point of breakdown for a synthetic social inclusion indicator is, however, arguable.[10] There are certain quality criteria for a synthetic indicator which only partially overlap with the quality criteria for a 'learning' set of sub-indicators. A marked difference between the two is the different stance towards highly intercorrelated sub-indicators. When the aim is to provide a better understanding of the causes and processes of social exclusion, high correlations may indeed be conducive to grasp what is essentially a

[9] Many of the data we use in our empirical application were taken from Eurostat's website list of structural indicators. If figures were unavailable there, we used data as given in the Joint Inclusion Report. Additional figures were taken from Eurostat's *Yearbook 2002*, and the Commission's publications *The social situation in the European Union* and *Social Protection in Europe.*

[10] In its 'Opinion on social indicators' of May 29, 2002, the Economic and Social Committee has emphasised that the indicators must be transparent and accessible to the public. We may therefore largely be dealing with a temporary problem.

multi-faceted phenomenon. Conversely though, this implies that there is little information lost in terms of *overall* explanatory power when one of two highly correlated indicators is neglected (see e.g. JRC (2002)). This observation is relevant to one of our excluded indicators, to wit, the rate of persistent low income. A statistic which measures how many people failed to attain the low income rate in at least two of three consecutive years, does tell us something about poverty dynamics. For the available data, it is strongly correlated with the low income rate (which *is* included in our analysis); the correlation coefficient amounts to 96.12%. This makes that excluding the rate of persistent low income hardly affects an analysis aimed at providing a summary picture: results would be roughly similar if the indicator were included, provided sufficient data were available.

However, this reasoning does not apply for the remaining two indicators on the list of structural indicators, viz. 'regional cohesion' and the 'number of people in jobless households'. The highest associated correlation with any of the 4 included indicators is 40.08% for the former and 43.27% for the latter (in both cases with the long term unemployment rate). For the jobless households indicator, the problem of limited data availability is probably of a temporary nature as EU members have committed themselves to regularly reporting this figure in the future. Still, it is a structural problem for the indicator of regional cohesion, which captures regional (coefficients of) variation in employment levels at Nuts2 level. For, given the absence of this particular regional breakdown-level, Denmark, Ireland and Luxemburg will never report the currently used regional cohesion figures. In section 3.3, we return to the issue when dealing with omitted performance dimensions.

For each indicator, we used the average of available data over the period 1995-2001 in our basic performance assessment model. Of course, this implies that our basic application is not a dynamic analysis. We take up a discussion of the possible dynamic use of the proposed methodology in section 3.2. The use of average data partly deals with the problem that data may be characterized by measurement errors; we can expect that these errors average out. Still, errors-in-the-data may partly influence our results. Again, we refer to our discussion below for other possible ways to deal with measurement errors in the current framework.

Table 3 displays the results of our basic, illustrative exercise. For each country, the *SI*-value and its associated optimal policy weights are displayed for two variants of model (2), viz. one in which we appended restriction (3), and another in which both restrictions (3) and (4) were included.

Table 3: *SI* results for EU countries*;* 2 weight configurations

	(each weight at least 10%)					(each weight at least 10%; at most 50%)				
	SI	**LIR weight**	**IQR weight**	**ESL weight**	**LTUR weight**	***SI***	**LIR weight**	**IQR weight**	**ESL weight**	**LTUR weight**
B	72.97%	10.00%	10.00%	70.00%	10.00%	68.12%	21.36%	10.00%	50.00%	18.64%
DK	100.00%	10.00%	43.74%	10.00%	36.26%	100.00%	10.00%	43.74%	10.00%	36.26%
D	76.52%	10.00%	10.00%	70.00%	10.00%	73.90%	10.00%	10.00%	50.00%	30.00%
EL	55.17%	10.00%	10.00%	70.00%	10.00%	51.96%	10.00%	10.00%	50.00%	30.00%
E	29.61%	10.00%	10.00%	70.00%	10.00%	26.60%	30.00%	10.00%	50.00%	10.00%
F	74.71%	10.00%	10.00%	70.00%	10.00%	72.00%	10.00%	10.00%	50.00%	30.00%
IRL	61.46%	10.00%	10.00%	51.16%	28.84%	61.33%	10.00%	10.00%	50.00%	30.00%
I	37.05%	10.00%	10.00%	70.00%	10.00%	35.92%	10.00%	30.00%	50.00%	10.00%
L	96.58%	10.00%	10.00%	10.00%	70.00%	90.94%	30.00%	10.00%	10.00%	50.00%
NL	89.62%	10.00%	10.00%	10.00%	70.00%	87.76%	30.00%	10.00%	10.00%	50.00%
A	98.40%	10.00%	10.00%	10.00%	70.00%	97.15%	10.00%	10.00%	30.00%	50.00%
P	62.02%	10.00%	10.00%	10.00%	70.00%	44.77%	10.00%	10.00%	30.00%	50.00%
FIN	100.00%	70.00%	10.00%	10.00%	10.00%	100.00%	50.00%	10.00%	30.00%	10.00%
S	100.00%	10.00%	10.00%	65.86%	14.14%	100.00%	23.19%	10.00%	50.00%	16.81%
UK	89.81%	10.00%	10.00%	70.00%	10.00%	88.84%	10.00%	10.00%	50.00%	30.00%

Notes: (1) B= Belgium; DK= Denmark; D= Germany; EL= Greece; E= Spain; F= France; IRL= Ireland; I= Italy; L= Luxembourg; NL= The Netherlands; A= Austria; P= Portugal; FIN= Finland; S= Sweden; UK= United Kingdom. (2) LIR, IQR, ESL and LTUR stand for (the weights accorded to) the low income rate, the income quintile ratio, early school leavers and long-term unemployment, respectively.

Apart from their strict interpretation as benefit-of-the-doubt weighted summary indices for social inclusion, these results suggest that a synthetic performance measure can be helpful in identifying patterns in the relative *global* performance over the EU countries. One notices e.g. the apparent difference between the southern and northern countries in our sample, or the intermediate position taken by the 'corporatist-conservative welfare state' cluster Germany, France and Belgium. One also observes that the majority of *SI* values are hardly affected by adding an extra (maximum) weight restriction in the second variant. The correlation between the two arrays of *SI* values amounts to more than 98.5% (as only Portugal's *SI* value drops considerably).

The tabulated weights are, to recall, endogenously generated by the maximization procedure and can thus reveal interesting information about the implicit (i.e. 'revealed') policy priorities. For countries that achieve a maximal *SI* value of 100%, the presented weights are not necessarily unique: there may exist other weight combinations that also yield a maximal *SI* value. The technical reason is that the underlying maximization procedure stops as soon as it reaches the maximum score of 100%. Which particular weight combination is selected then depends on the algorithm that is followed. Such multiple optimal solutions do generically not occur for countries with an *SI* value below 100%.

Nevertheless, even if some countries' tabulated weights characterize only a single optimum of the programming problem, the weights do possess useful indicative value. In Table 3, we observe that there appear to be important differences in the distribution of the weights over countries. For example, Germany, France, Belgium are among the countries which attach a high implicit weight to the educational dimension (see the maximal weight of 70% (50%) for the indicator 'early school leavers' in the left (right) half of Table 3). By contrast, other countries such as the Netherlands and Portugal, by virtue of the fact that their long term unemployment performance figures constitute their areas of comparative advantage, are taken to prioritize that dimension.

To conclude the discussion of our basic application, we point out that while the *SI* results are barely influenced by adding the maximal weight restriction, the revealed policy weights themselves seem to be a bit more responsive; weight correlations among the two variants range from 0.72 (low income rate) to 0.90 (early school leavers). Actually, in most cases covered by the left part of Table 3, an evaluated country assigns all (free) weight to just one dimension. Such phenomena are quite natural given that one only appends a single minimum weight restriction such as (3) to a benefit of the doubt weighting exercise. It further stands to reason that this phenomenon is mitigated as more (external) restrictions are added. At the right half of Table 3, countries have effectively been 'enforced' to further specify their implicit priorities. In that sense, the reported weights at the right side reveal more information than those with minimal (imposed) structure generated by the first variant. Again, we stress that, *globally,* only Portugal is really affected by the fact that it can no longer assign all free weight to its unemployment figure but must spread its remaining policy 'mass' over at least one other dimension.

3.2 Intertemporal Analysis

The proposed methodology can also be applied for intertemporal analysis. Specifically, it can be used to identify whether there is progress or regress in terms of social inclusion. We illustrate this for the years 1996 and 1997. The analysis concentrates on 14 EU countries only (due to limited data availability, we were not able to compute results for Luxembourg). Because of the data problems discussed before, this example again mainly serves illustrative purposes, and the results should be interpreted with caution.

Figure 1 depicts the *SI* results for the years 1996 and 1997. Performance results have been computed for the pooled country sample, i.e. including both the 1996 and 1997 observations.[11] The countries have been ranked according to their relative performance in 1996. For that year, we roughly get a similar picture as in Table 3 (which was based on 1995-2000 averages). Spain, Italy and Greece have rather low *SI* values, while Denmark, Finland Austria and Sweden are on top. We further find that most countries achieve a higher *SI* value in 1997 than in 1996; only Greece, Portugal and the UK exhibit a performance deterioration. Actually, the 1997 subsample includes the three 'benchmark' observations Finland, Denmark and Sweden (i.e. the observations associated with a 100% *SI* value).

These figures suggest a general social inclusion performance improvement within the EU between 1996 and 1997. This overall picture is confirmed by Figure 2. The performance deterioration of Greece, Portugal and the UK is rather small; it only marginally exceeds the 2% level. By contrast, the relative performance improvements are sometimes quite pronounced. For example, the relative performance of Spain and the Netherlands increased with almost 6% and 8%, respectively.

As this example makes clear, the possibilities of the proposed methodology are in principle not limited to computing *levels* of social inclusion performance: the same methodology is readily applicable for identifying *changes* in the relative performance. Evidently, such analysis of performance shifts over time can provide a useful input within the context of policy evaluations.

---

[11] See e.g. Färe and Grosskopf (1992) and Färe et al. (1994) for alternative procedures.

Figure 1: SI values for the years 1996 and 1997

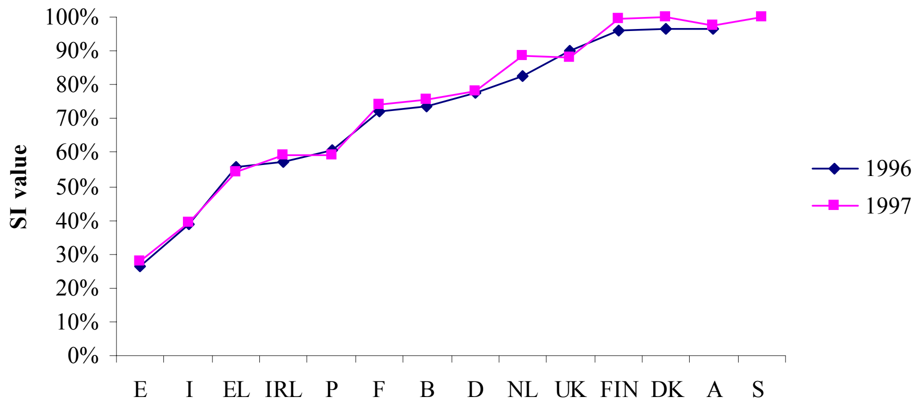
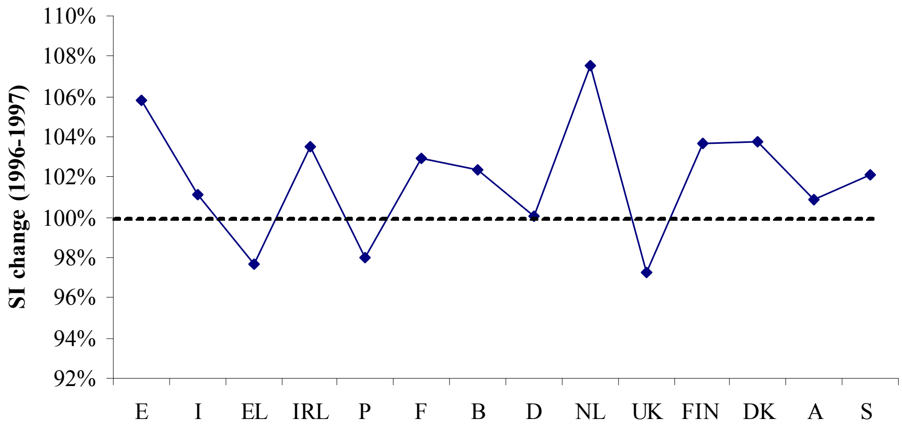


Figure 2: SI changes over 1996-1997



3.3 Sensitivity Analysis

Our previous discussion makes clear that, certainly in the current context, practical implementation of the proposed evaluation procedure almost inevitably entails a number of implicit, possibly debatable assumptions. In this section, we illustrate how one can check sensitivity of the performance results with respect to the weighting scheme and the number of original indicators within the proposed framework. Specifically, we examine the robustness of our basic analysis presented in section 3.1 (i.e., including 4 basic social performance dimensions, and using only a lower weight bound of 10%). We will refer to the associated performance index as the 'Basic' index.

OTHER WEIGHT ASSUMPTIONS

The endogenous determination of implicit policy weights constitutes the basic feature of the adopted methodology. In the construction of our Basic index, we imposed that each performance dimension should get a policy weight of at least 10%. The ad hoc nature of that assumption calls for investigating the sensitivity of our results. We have already discussed the impact of adding a maximal weight restriction in section 3.1. The general issue of sensitivity of a synthetic indicator to (pre-assigned) weight selection is also taken up in the JRC-report (2002), where interesting (resampling) procedures are discussed. Here, we confine ourselves to a comparison of our Basic index results with those of two alternative performance indexes: the index that imposes equal weights for all social inclusion indicators (EW) and the index that imposes no weight restrictions apart from non-negativity (NWR); see section 2.2 for the construction of these two alternative indexes. In a certain sense, these two alternatives can be considered as limiting cases in terms of the underlying (*a priori*) assumptions regarding the distribution of policy priorities.

A natural and concise way to compare the results associated with different performance indexes consists in looking at the associated correlation coefficients as in Table 4. The results in this table clearly illustrate the intermediate position of the Basic index; it correlates stronger with the EW index than the NWR index and stronger with the NWR index than the EW index. Further, we find that the Basic index correlates stronger with the NWR index than with the EW index. Overall, however, the relatively high correlation values suggest that different *a priori* assumptions regarding the distribution of policy priorities need not entail very divergent results on aggregate social inclusion performance. In other words, in the aggregate the specific weight restriction does not appear to have great impact on the calculated performance results. Recall, however, that models which endogenously determine the policy weights have clear value added in terms of 'revealing' implicit policy weights.

Although the aggregate impact of alternative weight restrictions may admittedly be rather limited, differences can be quite substantial at the level of individual countries. Indeed, for a number of countries we observe considerably different *SI* results dependent on the model that is used for quantifying overall performance (the Basic, the EW or the NWR variant); see Appendix 2. Yet in fact, these country-specific performance shifts can reveal interesting information. A notable example in this respect concerns the case of the UK. Looking at the original indicator values in Appendix 1, we find that this country achieves the lowest ESL value in our sample of 15 EU countries.[12] By contrast, the UK exhibits rather poor performance in terms of the LIR and IQR. Finally, the country performs moderately well in terms of the LTUR. Generally, this pattern suggests that the UK 'specializes' mainly in the educational dimension of social inclusion, and to a somewhat lesser extent in the employment dimension. The UK performance pattern is considerably different from that of, e.g., the Nordic countries, which perform relatively well in all 4 *SI* performance dimensions.

The EW aggregate performance index strongly penalizes this apparent specialization of the UK; the associated performance value amounts to only 62.10%. By contrast, the NWR index fully rewards such emphasizing of a single *SI* performance dimension; in this model, the UK achieves the maximum performance value of 100%. It could be argued, however, that such an index, completely determined by a sole performance dimension, fails in capturing *overall* social inclusion performance. By restricting the policy weights to be at least 10%, the Basic index, which equals 89.81% for the UK, corrects for such apparent 'over'-compensation. In fact, the only three countries that achieve the maximum performance value for this index are the 'well-balanced' Nordic countries.

[12] We note that this result may at least partly be due to data atypicity: the ESL figures we used for the UK were those appearing in the Annex of the Joint Inclusion Report since contrary to other countries, there are currently no ESL-figures displayed for the UK in Eurostat's website list of structural indicators.

This example suggests that the (Basic) weight-restricted *SI* model conveniently combines the attractive features of the 'extreme' EW and NWR models: (1) it allows for assigning national policy priorities (unlike the EW model); and yet (2) it penalizes over-specialization in a single or only a few policy dimensions (unlike the NWR model).

DIMENSIONS

Another aspect of sensitivity analysis pertains to the selection of the performance dimensions. As discussed in section 3.1, limited data availability prevented us from including all official indicators of social inclusion in the Basic index. This suggests examining the robustness of the performance results with respect to the omitted performance dimensions. In particular, we consider 'jobless households' and 'regional cohesion', which do not strongly correlate with the 4 indicators that are included in the basic model. In addition, we consider the impact of other indicators on the social scoreboard, viz. those pertaining to the health status of a population (as reflected by 'life expectancy at birth' and the 'health' indicator, which captures the self-perceived health status by income level). The original (averaged) data are reported in Appendix 1.

As before, the impact of omitting a particular performance dimension can be checked by comparing (in terms of correlation coefficients) the results of the Basic index, which does not include the corresponding indicator, with those of the (similarly constructed) index that does incorporate the indicator. In some cases, this entails comparison of samples of different size. For example, regional cohesion and jobless household values are reported for only 12 EU countries, and health values are available for only 13 EU countries. Evidently, in those cases we compute index values and correlation coefficients for evaluation models that include the smaller samples.

For our application, we find that omitting the aforementioned performance dimensions has little impact on the performance results (see again Table 4, column 1): the correlation between the newly constructed indexes and the Basic index everywhere amounts to more than 90%. This suggests that excluding these dimensions from the social inclusion analysis does not have strong distortionary effects: good (poor) performers generally remain good (poor) performers.

Again, it may be informative to consider performance shifts at the level of individual countries; see the results in Appendix 2. In that respect, the substantial increase of the Portuguese index value when including jobless households in the analysis is especially remarkable. Indeed, this country achieves the best performance in that social inclusion dimension, while performance in the dimensions that underlie the Basic index (except from the long term unemployment) is generally very poor; see Appendix 1. Hence, including the jobless household dimension in the computation of the (unequally weighted) performance index naturally increases Portugal's associated *SI* value.

COMPARISON WITH OTHER SOCIAL PERFORMANCE INDICATORS

As a final validity check, we compare our results with those of alternative measures that have been used for cross-country assessments of social policies. In particular, we look at the Human Development Index (HDI) and the Human Poverty Index (HPI) (UNDP, 2001), suggested as measures for the economic and social aspects of sustainable development. We also consider the PAIN index, which Asher et al. (1993) proposed as a summary measure of distributive justice. Our specific construction of the PAIN-index is different from that originally proposed (it was originally constructed with fixed pre-assigned weights); we aggregate the poverty rate and the Gini coefficient using the *SI* model presented in section 2.2 (including a lower weight bound of 10%).

The relevant correlation coefficients are presented in the bottom rows of Table 4. Not very surprisingly, we find a positive correlation between the Basic index and each of the 3 alternative social performance indexes. The strongest correlation is associated with the PAIN index; this index indeed being most directly comparable in terms of its specific scope. Next, we find that our Basic *SI* correlates stronger with the HDI than with the HPI. This may appear somewhat counter-intuitive, given that the HPI comes closer in spirit to the social inclusion concept than the HDI. Still, the differences between the HDI and HPI correlation coefficients remain rather small.

The correlation is far from perfect. Of course, this reflects the different construction of the alternative indexes, both in terms of the dimensions that are included and (only for the HPI and HDI) in terms of the underlying weighting procedure. On the other hand, it also indicates that a full assessment of social policy can benefit from comparing the results for differently conceptualized performance indexes: dependent on its specific orientation, each index contains useful information concerning particular policy aspects.

Table 4: alternative synthetic indicators; correlation matrix

	Basic	EW	NWR	+ Regcoh	+ Jl H	+ Health	+ LE	HDI	HPI	PAIN
Basic	100.00%									
EW	92.98%	100.00%								
NWR	97.88%	83.81%	100.00%							
+ Regcoh	90.58%	85.92%	87.63%	100.00%						
+ Jl H	90.72%	71.76%	94.70%	82.26%	100.00%					
+ Health	97.36%	88.42%	95.43%	86.85%	82.41%	100.00%				
+ LE	94.96%	96.11%	88.25%	87.25%	80.28%	88.91%	100.00%			
HDI	56.23%	74.25%	43.18%	43.07%	13.92%	51.35%	62.72%	100.00%		
HPI	47.40%	63.30%	38.83%	49.13%	30.27%	12.38%	60.34%	49.52%	100.00%	
PAIN	77.64%	93.27%	64.53%	70.52%	49.67%	64.80%	88.43%	74.44%	82.41%	100.00%

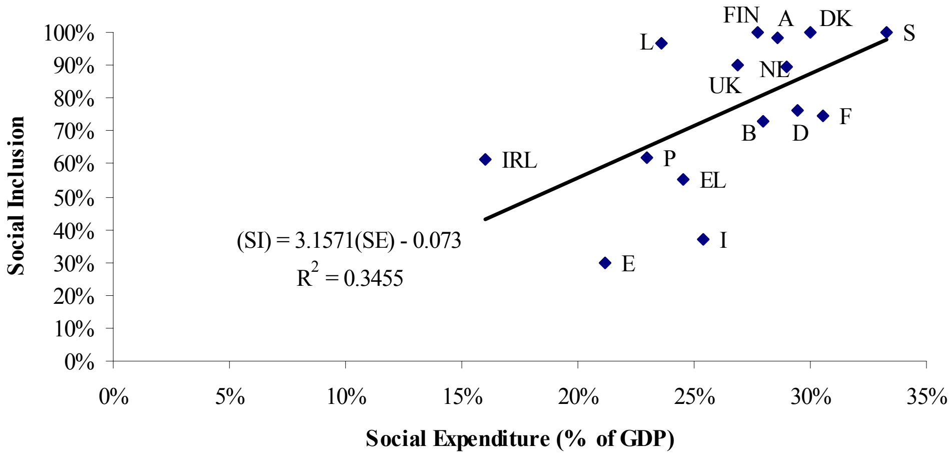
Note: Basic stands for the basic configuration; EW stands for basic selection of dimensions with equal weights; NWR stands for basic selection of dimensions and no weight restrictions; +Regcoh stands for basic with 'regional cohesion' included; + Jl H stands for basic with 'jobless households' included; + Health stands for basic configuration with 'Health" included; +LE stands for basic configuration with 'Life expectancy' included; HDI stands for 'Human Development Index'; HPI stands for 'Human Poverty Index'; PAIN stands for 'PAIN index'.

3.4 Performance analysis versus efficiency analysis

As both the scoreboard and the index (deliberately) focus solely on policy outcomes, they can surely be criticized precisely because they totally abstract from the inputs that nations employ for their social inclusion goals. Yet in all likelihood, advancing social inclusion is not a free lunch. For example, there is quite some statistical evidence that there exists a negative relationship between a country's social spending/GDP ratio and its low income rate (see e.g. Gouyette and Pestieau, 1999). Can a similar relationship be discerned between total financial inputs and an index which captures all (policy-weighted) outcomes simultaneously?

Figure 3 summarises elementary results from such an exercise. We regressed the Basic index values on (gross) social expenditure as a % of GDP. Although the particular relationship may well be positive but nonlinear, we have added a linear regression line together with its goodness-of-fit statistic (R^2=0.3455). Interestingly, this implies a moderately better goodness-of-fit in comparison with a similar (simple) linear regression which only considers the low income rate as dependent variable.[13]

Figure 3: Social expenditure and the Basic *SI*-index



Of course, once that policy inputs are allowed to enter the picture, it may well be that a mediocre outcome is achieved economically, just as it may be that impressive performance is accomplished inefficiently (indeed, national politicians do face more trade-offs than those related to social inclusion). We will not develop the issue of efficiency measurement in this paper, but confine ourselves to recalling that the method we discuss is in point of fact rooted in the theory of productivity measurement. Hence, the above procedure can be quite naturally adapted to take inputs –and other outputs– into account (see Zaim et al. (2001) for such an application based on the UNDP-HDI), or even for genuinely assessing the relative productive efficiency of social service provision.

---

[13] For our data, we have (LIR) = 0.3142 – 0.5915 (SE), and R^2 = 0.3069. Actually, the fact that the use of a synthetic index as dependent variable may increase the overall fit has also been observed in the context of synthetic indices for macroeconomic performance (versus the single dimensional GDP-growth variable), when investigating their relationship with social capital. See Moesen et al. (1998).

4. CONCLUDING REMARKS

In this paper we focused on a set of indicators for social inclusion in its capacity as a social scoreboard. We presented a methodology for merging the various dimensions into one synthetic indicator of social inclusion. The obtained single-valued index has an intuitive, easy to convey interpretation and is receptive to valuing the legitimate diversity of national policies. The methodology, grounded in microeconomic theory, is based on the benefit-of-the-doubt principle, which allows countries to attach higher weights to those social inclusion dimensions in which they perform relatively well. The endogenously defined weights can be interpreted as implicitly revealed policy priorities. Additionally, the method straightforwardly allows for incorporating all kinds of prior assumptions concerning the feasible range of policy weights. And, the method can readily include external targets or external comparison partners in social inclusion performance assessments.

We have illustrated the methodology by means of a basic application to the 15 EU countries, assessing social inclusion performance for the period 1995-2001. Subsequently, we have shown how one can use the proposed method for intertemporal performance assessments, and have indicated how one can examine the sensitivity of the obtained performance results with respect to the policy weighting assumptions that are used and the social inclusion dimensions that are included. Finally, we have touched upon the relationship between synthetic social inclusion policy performance and social inclusion policy efficiency.

For the sake of transparency, we have mainly limited our discussion to introducing the basic methodology, and opted for a very basic application which was principally intended to serve illustrative purposes. This makes that we have not fully addressed a number of issues that become of essence in more extensive practical applications:

1. First, we have largely left open the issue of specifying prior assumptions concerning the distribution of policy priorities. In our application, we used minimal assumptions concerning generally acceptable lower and upper bounds for the relative weights, and we complemented this with a general discussion regarding the possibility of checking sensitivity with respect to the employed weighting assumptions. As we have repeatedly stressed, however, additional (more stringent) weight bounds can readily be incorporated in the evaluation models. Such assumptions can further benefit the analysis. They may, e.g., reflect expert consensus positions. Generally, we believe that the conception of such additional weight restrictions constitutes a promising avenue for further research.

2. Next, we have chosen to refrain from an explicit discussion of dealing with low quality data. For example, we have not explicitly addressed the issue of measurement errors, which is certainly important in today's practical context of social inclusion assessment in the EU. In our application, we simply employed average data (over the period 1995-2001) to partly deal with that problem, as it can be expected that errors-in-the-data average out. Still, more advanced techniques to deal with practical data problems have been proposed in the closely related productivity measurement (or 'Data Envelopment Analysis') literature; see Cherchye and Post (2001) for an up-to-date discussion. We confine ourselves here to stating that these techniques can be readily integrated in the presented framework.

3. More generally, we have limited our empirical analysis to 'point estimates' for the *SI* values and policy weights. Of course, it would be interesting to complement such point

estimates with confidence intervals. These can be constructed by using resampling techniques (see Simar and Wilson, 1998). Normally, point estimates will be 'more precise' as data sets get larger. Addition of the acceding countries' indicators to the OMC, but also our earlier suggestion to enlarge the set of possible comparison partners with external, i.e. non EU observations, will in that sense provide a *clearer* big picture. Within this line of reasoning, we point out there is also a trade-off between the number of (single indicator) dimensions considered for the construction of a synthetic indicator and the asymptotic efficiency of the estimates, a problem that is known as 'the curse of dimensionality' in nonparametric estimation models; within the nonparametric productivity analysis literature, Simar and Wilson (2000) provide a good methodological survey of statistical testing possibilities.

4. Finally, we have addressed the issue of variable selection by means of an easily implemented, but rather superficial correlation comparison analysis. We again refer to the well-established body of literature on linear programming approaches to productivity measurement for more refined, statistical-based variables selection techniques; see, e.g., Pastor et al. (2002) for a recent proposal. These can equally be employed in the presented framework. Another promising avenue consists in reducing the original number of social inclusion indicators by means of principal components analysis techniques; see, e.g., Adler and Golany (2001) for an application within a Data Envelopment Analysis setting.

While the Commission has recently affirmed its willingness to use synthetic indicators to complement existing single-dimensional gauges, it is at this stage really an open question whether their application will ever be deemed appropriate (not in the least by national decision makers) in the context of the OMC for poverty and social inclusion. We are, however, not alone in believing that there is a case for increasing public awareness about the method, and we do think that a well-designed and robust synthetic indicator may for that reason suit OMC's ultimate purpose.

REFERENCES

Adler, N., B. Golany (2001), "Evaluation of Deregulated Airline Networks Using Data Envelopment Analysis Combined with Principal Component Analysis; with an Application to Western Europe", *European Journal of Operational Research* **132**, 260-273.

Asher,M., R. Defina, K. Thanawalla (1993), "The Misery Index: Only Part of the Story", *Challenge*, 58-62.

Atkinson, A., B. Cantillon, E. Marlier, B. Nolan (2002), *Social Indicators. The EU and Social Inclusion*, Oxford University Press.

Atkinson, A. (2002a), *Developing Comparable Indicators for Monitoring Social Inclusion in the European Union*, mimeo.

Atkinson, A. (2002b), "Social Inclusion and the European Union", *Journal of Common Market Studies* **40**, 625-643.

Balk, B. (2002), "*The Residual. On Monitoring and Benchmarking Firms, Industries and Economies with respect to Productivity*", Erasmus Research Institute of Management, EIA-2002-07-MKT.

Brandolini, A. (2002), "Education and Employment Indicators for the EU Social Agenda", *Politica Economica* **18**, 55-62.

Charnes, A., W.W. Cooper, E. Rhodes (1978), "Measuring the Efficiency of Decision Making Units", *European Journal of Operations Research* **2**, 429-444.

Cherchye, L. (2001), "Using Data Envelopment Analysis to Assess Macroeconomic Policy Performance", *Applied Economics* **33**, 407-416.

Cherchye, L., T. Kuosmanen (2002), "*Benchmarking Sustainable Development: A Synthetic Meta-Index Approach*", EconWPA Working Paper.

Cherchye, L., T. Post (2001), "Methodological Advances in DEA: A survey and an application for the Dutch electricity sector", Erasmus Research Institute of Management ERS-2001-53-F&A, forthcoming in *Statistica Neerlandica*.

Duffy, K. (2002), *EAPN follow up on the National Action Plans on Inclusion: The State of Play by June 2002*, EAPN Report.

Färe, R., S. Grosskopf (1992), "Malmquist Productivity Indices and Fisher Ideal Indices", *Economic Journal* **102**, 158-160

Färe, R., S. Grosskopf, M. Norris, and Z. Zhang (1994), "Productivity Growth, Technical Progress and Efficiency Change in Industrialised Countries", *American Economic Review* **84**, 66-83.

European Commission (2002), '*Structural Indicators*', Communication from the Commission, COM (2002), 551 Final.

Giammusso, F., R. Tangorra (2002), "Indicators on Poverty and Social Exclusion: Moving Toward a Social Europe?", *Politica Economica* **18**, 29-47.

Gouyette, C., P. Pestieau (1999), "Efficiency of the Welfare State", *Kyklos* **52**, 537-553.

Hemerijck, A. (2002), "The Self-Transformation of the European Social Models", in G. Esping-Andersen e.a., Why we need a New Welfare State, Oxford University Press, 173-213.

Hills, J. (2002), "Comprehensibility and Balance: The Case for Putting Indicators in Baskets", *Politica Economica* **18**, 95-98.

Joint Research Centre of the European Commission-Applied Statistics Centre (2002), *State-of-the-art Report on Current Methodologies and Practices for Composite Indicator Development*, ISPRA, 72 p.

Lovell, C.A.K., Pastor, J.T. (1995), "Units Invariant and Translation Invariant DEA Models", *Operations Research Letters* **18**, 147-151.

Mahlberg, B., M. Obersteiner (2001), *Remeasuring the HDI by Data Envelopment Analysis*, International Institute for Applied Systems Analysis Interim Report 01-069.

Melyn, W., W. Moesen (1991), *Towards a Synthetic Indicator of Macroeconomic Performance: Unequal Weighting when Limited Information is Available*, Public Economics Research paper 17, CES, KU Leuven.

Micklewright, J. (2001), "Should the UK Government Measure Poverty and Social Exclusion with a Composite Index?' in *CASE, Indicators of Progress: A Discussion of Approaches to Monitor the Government's Strategy to Tackle Poverty and Social Exclusion*, CASE Report 13, London School of Economics.

Moesen, W., L. Cherchye, T. Van Puyenbroeck (1998), "*It Pays to Be Decent: on the relationship between Microsocietal Values and the Macroeconomic Performance of Nations*", mimeo.

Pastor, J.T., J.L. Ruiz, I. Sirvent (2002), "A Statistical Test for Nested DEA Models", *Operations Research* **50**, 728-735.

Pedraja-Chaporro, F., J. Salinas-Jimenez, P.Smith (1997), "On the Role of Weight Restrictions in Data Envelopment Analysis", *Journal of Productivity Analysis* **8**, 215-230.

Scharpf, F. (2002a), *Legitimate Diversity: the New Challenge of European Integration*, Cahiers Européens de Sciences PO, 1/2002.

Scharpf, F. (2002b), "The European Social Model", *Journal of Common Market Studies* **40**, 645-670.

Simar, L., P. Wilson (1998), “Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Non-Parametric Frontier Models”, *Management Science* **44**, 49-61.

Simar, L., P. Wilson (2000), “Statistical Inference in Nonparametric Frontier Models: the State of the Art”, *Journal of Productivity Analysis* **13**, 49-78.

Srinivasan, T.N. (1994), “Human Development: A New Paradigm or Reinvention of the Wheel?, *American Economic Review* **84** (2), 238-242.

Storrie, D., H. Bjurek (2000), *Benchmarking European Labour Market Performance with Efficiency Frontier Techniques*, CELMS Discussion paper, Göteborg University.

UNDP (2001): *Human Development Report 2001: Making new technologies work for human development*, Oxford University Press, Oxford, New York.

Vandenbroucke, F. (2002), “Foreword”, in Atkinson et al., *Social Indicators. The EU and Social Inclusion*, Oxford University Press, v-xi.

Visser, J., A. Hemerijck (2001), *Learning Ahead of Failure: A Research Agenda*, University of Amsterdam.

Zaim, O., R. Färe, S. Grosskopf (2001), “An Economic Approach to Achievement and Improvement Indexes”, *Social Indicators Research* **56**, 91-118.

APPENDIX 1: DATA; AVERAGE VALUES 1995-2001

	LIR	IQR	ESL	LTUR	Regcoh	Jl H	H	LE	G	SE
B	16.000	5.900	13.786	5.217	45.771	17.871	0.273	74.000	35.000	27.933
DK	9.750	2.775	11.229	1.529			0.545	73.500	21.667	29.967
D	15.750	5.025	13.700	4.417	44.643	15.400	0.600	74.020	30.000	29.467
EL	22.000	6.550	19.171	5.571	21.143	13.057	0.267	75.340	34.667	24.533
E	19.500	6.675	30.100	8.929	31.271	13.150	0.429	74.800	34.667	21.200
F	16.750	4.750	14.417	4.583	21.843	15.233	0.364	74.460	29.667	30.533
IRL	19.000	5.550	19.550	4.343		15.533	0.250	73.360	33.667	16.000
I	19.500	5.900	28.700	6.929	65.300	15.071	0.438	75.240	32.333	25.400
L	12.000	4.700	25.567	0.750		10.286	0.133	73.760	28.500	23.600
NL	11.500	4.550	16.017	1.871	19.200	12.343	0.429	75.060	29.333	28.967
A	13.250	4.025	11.267	1.086	26.886	10.843	0.308	74.320	26.333	28.600
P	22.250	7.275	42.650	2.357	29.386	7.333	0.300	71.520	37.667	22.933
FIN	8.000	2.900	9.367	4.014	22.771			73.300	22.500	27.733
S	11.000	3.467	7.225	2.429	20.700			76.680	23.000	33.267
UK	20.250	5.900	6.000	2.271	40.257	16.686	0.167	74.560	34.000	26.900

Note: LIR stands for 'low income rate' (60% level); IGR stands for 'income quintile ratio'; ESL stands for 'early school leavers'; LTUR stands for 'long-term unemployment rate'; Regcoh stands for 'Regional Cohesion'; Jl H stands for 'Jobless Households'; H stands for 'Health'; LE stands for 'Life Expectancy'; G stands for 'Gini coefficient'. See footnote 10 for sources.

APPENDIX 2: PERFORMANCE RESULTS; ALTERNATIVE CONFIGURATIONS

	Basic	EW	NWR	+ Regcoh	+ Jl H	+ H	+ LE	HDI	HPI	PAIN
B	72.97%	54.56%	80.43%	69.01%	73.38%	83.41%	69.38%	100.00%	94.00%	41.36%
DK	100.00%	100.00%	100.00%			100.00%	100.00%	100.00%	98.00%	100.00%
D	76.52%	63.14%	80.76%	72.80%	79.85%	76.96%	72.93%	98.00%	98.00%	49.02%
EL	55.17%	33.79%	64.06%	75.82%	60.24%	68.04%	60.35%	94.00%		17.26%
E	29.61%	18.38%	34.98%	55.40%	46.81%	36.32%	47.70%	97.00%	95.00%	19.63%
F	74.71%	61.79%	79.01%	85.90%	78.24%	81.49%	72.27%	98.00%	95.00%	49.47%
IRL	61.46%	49.53%	66.22%		63.76%	77.72%	58.98%	98.00%	91.00%	25.27%
I	37.05%	30.88%	39.21%	32.91%	40.56%	43.57%	57.99%	97.00%	94.00%	32.33%
L	96.58%	75.77%	100.00%		100.00%	100.00%	95.84%	98.00%	96.00%	70.83%
NL	89.62%	81.05%	92.76%	98.50%	100.00%	90.97%	93.90%	99.00%	98.00%	73.49%
A	98.40%	87.08%	100.00%	100.00%	100.00%	100.00%	99.82%	98.00%		71.15%
P	62.02%	22.08%	80.35%	64.65%	94.88%	62.39%	56.40%	93.00%		0.00%
FIN	100.00%	95.66%	100.00%	100.00%			100.00%	99.00%	98.00%	100.00%
S	100.00%	93.34%	100.00%	100.00%			100.00%	100.00%	100.00%	91.52%
UK	89.81%	62.10%	100.00%	84.84%	90.82%	100.00%	85.04%	98.00%	91.00%	22.30%

Note: Basic stands for the basic configuration; EW stands for basic selection of dimensions with equal weights; NWR stands for basic selection of dimensions and no weight restrictions; +Regcoh stands for basic with 'regional cohesion' included; + Jl H stands for basic with 'jobless households' included; + H stands for basic configuration with 'Health' included; +LE stands for basic configuration with 'Life expectancy' included; HDI stands for 'Human Development Index'; HPI stands for 'Human Poverty Index'; PAIN stands for 'PAIN index'.